DETECTION OF MOTION USING THE WAVELET TRANSFORM

A Thesis Submitted

in Partial Fulfillment of the Requirements

for the Degree of

Master of Technology

by

Capt. Deepak Sharma

to the

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

April 1994

Certificate

It is certified that the work contained in the thesis entitled **DETECTION OF MOTION USING THE WAVELET TRANSFORM,** by Capt. Deepak Sharma, has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

April 1994

Dr. Sumana Gupta
Associate Professor
Department of Electrical Engineering
I.I.T. Kanpur

Abstract

A method for the detection and estimation of motion in image sequences is presented.In order to estimate motion parameters reliably, a gradient based spatio-temporal constrain equation for motion estimation using zero crossings of wavelet transform is described. Multiresolution image decomposition is performed with the biorthogonal wavelet transform and motion parameters are hierarchially estimated.Motion vectors are also estimated using Laplacian of Gaussian smoothing filter,using the same constrain equation.Finally a performance comparasion of the two methods are carried out using the synthetic and laboratory image sequences respectively.

To

My Parents

Acknowledgements

I would like to convey my sincere gratitude to Dr Sumana Gupta who has extended all the guidance and necessary advice at all the times to make this work a success. She has been a constant source of encouragement during the work.

I am also thankful to Lieutenant M.D.Patil who extended the necessary help at the crucial moment.

Deepak Sharma

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Velocity information is important not only for detecting velocities and trajectories of objects but also as a clue for image segmentation. It may be used to create new frames from the existing one's through temporal interpolation. Therefore a new frame is usually created from two adjacent frames one in the past and one in future relative to the frame being created. A straight-forward approach for obtaining velocity information is to find the prominent features and track them by a matching method from frame to frame [1]. The disadvantage with this method is that the computations required are very large and also if the initial velocity vector is wrongly estimated, it leads to an erroneous final velocity estimation.

Another approach [2] uses the fact that for a moving objects the spatial and temporal intensity changes are not independent of each other and are given by the relation:

$$V_x \frac{\partial I(x,y,t)}{\partial x} + V_y \frac{\partial I(x,y,t)}{\partial y} + \frac{\partial I(x,y,t)}{\partial t} = 0 \tag{1.1}$$

It is assumed that the image intensities are differentiable; V_x and V_y are the two components of the velocity vector; (dI/dx and dI/dy) are the two components of the spatial and temporal intensity gradients; (dI/dt) represents the gray level difference

between consecutive frames at a point (x,y) [2].

Variation of the basic method [2] involves processing of the image before subtraction. For instance in the basic detection scheme [3, 4] each image intensity ' I ' is first convolved with the Laplacian of the Gaussian smoothing function $\nabla^2 G(x, y)$ and the zero crossings in the single frame of $\nabla^2 G(x, y) * I$ are located.The change in the value of $\nabla G(x, y) * I$ from one frame to the next is non zero only when the edges have moved. This method can detect motion vectors correctly to some extent if the pixel displacement per frame is small ie of the order of one or two pixel per frame. The method however fails for large values of pixel displacement per frame.

In this thesis we propose a variation of subtraction method based on spatio-temporal constrain equation [3, 4]. This hierarchial method for estimating motion parameters uses zero crossings of wavelet transform [5]. A multiresolution image decomposition is performed using biorthogonal wavelet transform and the motion is hierarchially estimated. It is shown that the motion parameters estimated are reasonably accurate and the method works satisfactorily for values of pixel displacement greater than 10 pixels per frame. It is also shown that the spatial and temporal derivatives of the function $\nabla^2 G(x, y) * I$ can be used to compute the component of velocity normal to the zero crossing contours. The computation of the normal component of the velocity vector is quite similar to other methods including the work of Marr and Ullman [6].

1.1 Scene Segmentation

In an image sequence, if all the pixel in a frame are moving with the same set of motion parameters, then the correlation between two consecutive images is maximized when

the first image is transformed according to the motion parameters. This is the result of the well known fact that, given two images,their correlation is maximized when one of them is a replica of other even if scaled by a constant. The situation becomes complicated if there are more than one moving object. In the correlation method it is observed that multiple moving objects give rise to multiple peaks correspond to the velocity of moving objects. This observation can be used for detecting multiple moving moving objects.

Let us suppose that there are two moving objects in a frame, and the magnitude of individual motion is small This will contribute to two peaks. False peaks are possible if the magnitude of individual motion is large, resulting in aliasing, this can be avoided by low pass filtering of the image.Another possible source of false peaks is due to merging of peaks depending on the relative broadness of the peaks.If two peaks of roughly the same height are brought together sufficiently close, the pair of broad peaks bear a higher risk of merging together to form a single peak.For the same separation the risk is less for a pair of sharper peaks to merge together to form a single peak. The High-pass filtering for example, can be used to sharpen the correlation function.

Now we have two contradictory requirements. On one hand, low-pass filtering is needed to avoid false peaks and on other hand high-pass filtering is needed to avoid merging of peaks. Facing this dilemma, a possible solution is a multiresolution scheme. This problem can be well resolved using Wavelet Transform; which decomposes image into multiresolution level. Thus a suitable scene segmentation can be achieved using multiresolution decomposition of image, using the wavelet transform.

1.2 Problem Statement

In this thesis a new method for detection and estimation of motion with wavelet transform is proposed. The image sequences are decomposed into various resolution levels using biorthogonal wavelet transform and the motion vectors are hierarchially estimated using spatio-temporal constrain equation. The multiresolution decomposition of image is a pseudo frame; and the motion vectors estimated at the coarser layer are used to initialise the estimation of motion vectors at the next resolution level. The motion parameters are hierarchilly estimated.

A comparision of motion vectors evaluated using Laplacian of Gaussian smoothing function, for the same gradient constrain equation was made with the hierachial method. The comparision showed that the hierarchial method of estimating motion parameters has overcome the drawbacks of Laplacian of Gaussian method.

1.3 Organisation of Thesis

The thesis has been presented in the five chapters. The first chapter contains the introduction and the summary of the work. The second chapter deals with the general theory of estimating motion parameters. In the third chapter wavelet transform is briefly discussed. Chapter four discusses in detail the method of motion estimation using wavelet transform. The motion parameters were estimated using Laplacian of Gaussian smoothing function. The motion parameters evaluated using both the methods for small as well as for the large value of pixel displacement per frame were analysed and compared in each case. Chapter five concludes the thesis and discusses the scope for future work.

Chapter 2

MOTION ESTIMATION

2.1 Introduction

The process of determining the movement of objects within the sequence of image frames is known as motion estimation. Processing images accounting for the presence of motion, is called motion compensated image processing

Motion compensated image processing has variety of applications. One application is image interpolation. By estimating motion parameters we can create a new frame between two adjacent existing frames.

In motion estimation problem we consider here, only translational motion of objects.

2.2 Various Methods

Let $I(x, y, t_{-1})$ and $I(x, y, t_0)$ denote the image intensities at t_{-1} and t_0 respectively. We refer to $I(x, y, t_{-1})$ and $I(x, y, t_0)$ as past and current frame. We assume

$$I(x, y, t_0) = I(x - dx, y - dy, t_{-1}) \tag{2.1}$$

where dx and dy are horizontal and vertical displacement between t_{-1} and t_0.

$$I(x,y,t) = I((x - V_x(t - t_{-1}), y - V_y(t - t_{-1})) \quad t_{-1} \leq t \leq t_0 \tag{2.2}$$

A direct consequence of equation 2.2 is a differential equation which relates V_x and V_y to $\frac{\partial I(x,y,t)}{\partial x}, \frac{\partial I(x,y,t)}{\partial y}$ and $\frac{\partial I(x,y,t)}{\partial t}$ which is valid in the spatio-temporal region over which uniform motion is assumed. To derive the relationship. Let

$$s(x,y) = I(x,y,t_{-1}) \tag{2.3}$$

from equations 2.2 and 2.3

$$I(x,y,t) = s(a(x,y,t), b(x,y,t)) \quad t_{-1} \leq t \leq t_0 \tag{2.4}$$

where

$a(x,y,t) = x - V_x(t - t_{-1})$ and $b(x,y,t) = y - V_y(t - t_{-1})$

From above we obtain

$$\frac{\partial I(x,y,t)}{\partial x} = \frac{\partial s}{\partial a}\frac{\partial a}{\partial x} + \frac{\partial s}{\partial b}\frac{\partial b}{\partial x} = \frac{\partial s}{\partial a} \tag{2.5}$$

$$\frac{\partial I(x,y,t)}{\partial y} = \frac{\partial s}{\partial a}\frac{\partial a}{\partial y} + \frac{\partial s}{\partial b}\frac{\partial b}{\partial y} = \frac{\partial s}{\partial b} \tag{2.6}$$

$$\frac{\partial I(x,y,t)}{\partial t} = \frac{\partial s}{\partial a}\frac{\partial a}{\partial t} + \frac{\partial s}{\partial b}\frac{\partial b}{\partial t} \tag{2.7}$$

From 2.7

$$V_x\frac{\partial I(x,y,t)}{\partial x} + V_y\frac{\partial I(x,y,t)}{\partial y} + \frac{\partial I(x,y,t)}{\partial t} = 0 \tag{2.8}$$

Equation 2.8 is called the spatio-temporal constrain equation and can be generalised to incorporate other types of motions such as zooming etc. equations 2.1 and 2.8 are highly restrictive. For example they do not allow for object rotation, camera zoom, regions uncovered by translational motion and multiple objects moving with

different velocities. Here it is assumed that the background region is not affected by object motion and the object is moving with uniform translatory motion between two successive frames under consideration.

Motion estimation problem can be classified broadly into two groups.

1. Region matching methods.

2. spatio-temporal constraint methods.

Region matching methods are based on equation 2.1 ; and constraint methods are based on equation 2.8.

2.3 Region Matching Methods

Region matching methods involve considering a small region in a frame and searching for the displacement which produces the best match among possible regions in an adjacent frame. In region matching method the displacement vector $[dx, dy]$ is estimated by minimising:

$$Error = \int_x \int_y ((I(x,y,t_0) - I(x,y,t_{-1}))^2 dx, dy \quad (2.9)$$

$$or \;\; Error = \int_x \int_y \mid I(x,y,t_0) - I(x,y,t_{-1}) \mid dx, dy \quad (2.10)$$

The functions $I(x,y,t_0) - I(x,y,t_{-1})$ is called displaced frame difference.The error expression 2.9 and 2.10 is zero at correct displacement $[dx, dy]$. Minimising equation 2.9 and equation 2.10 is a nonlinear problem.Attempts to solve this nonlinear problem has produced many variations, which can be grouped into block matching and recursive methods. We discuss briefly block matching.

2.4 Block Matching method

One straight-forward approach to solve the above minimisation problem is to evaluate the error for every possible $[dx, dy]$ within some reasonable range and choose $[dx, dy]$ at which the error is minimum. In this method, a block of pixel intensities at t_0 is matched directly to a block at time t_{-1}. This is the basis of the block matching method. This method is computationally expensive and many methods have been developed to reduce computations. In the case of three step search method, in the first step, the error expression is evaluated at nine values of (dx,dy). Among these nine values we choose (dx,dy) with the smallest error. In second step, we evaluate error expression at eight additional values of (dx,dy). We now choose from the nine values, that is one previous and eight current values of (dx,dy), the minimum of (dx,dy). This procedure is repeated once more. At the end of third step, we have an estimate of (dx,dy). This procedure can be easily extended to more-than three steps to estimate (dx,dy). There are other methods to solve the region matching problem [1, 2].

2.5 Spatio-temporal Constrain Method

Algorithms of this class are based on the equation 2.8,which can be viewed as a linear equation of two unknown parameters (Vx, Vy) under the assumption that image intensities are differentiable.

In the simple subtraction methods,the intensities at each pixel at adjacent points in time are subtracted from each other; nonzero values in the resulting difference image frame indicate that something in the the image has changed. It is assumed that these changes are due to motion,rather than the illumination effects. Variation

of this involve processing of image before subtraction [3].

2.5.1 Laplacian of Gaussian Method

In this, the image intensity field 'I' is first convolved with Laplacian of Gaussian smoothing function $\nabla^2 G(x,y)$ and the zero crossing in the single frame $\nabla^2 G(x,y) * I$ are located. The change in the value of $\nabla^2 G(x,y) * I$ from one frame to the next at the location of zero crossings in the current frame are nonzero only if the edges have moved. If no motion has taken place and there is a change only in illumination, then the change in value of $\nabla^2 G(x,y) * I$ from one frame to the next at the location of zero crossings in the current frame will be zero. Thus this method can discriminate between motion and time varying illumination. Equation 2.8 can be used to compute the normal component of velocity vector [3]. The magnitude of the normal component of velocity 'Un' at the zero crossing is given by [3]:

$$U_n = -\frac{\frac{\partial I}{\partial t}}{\sqrt{(\frac{\partial I}{\partial x})^2 + (\frac{\partial I}{\partial y})^2}} \tag{2.11}$$

and the angle 'θ' between the unit normal to the contour and the positive x-axis is given by

$$\theta = tan^{-1}\frac{\frac{\partial I}{\partial y}}{\frac{\partial I}{\partial x}} \tag{2.12}$$

2.6 Image Interpolation

One of the application of motion estimation is in image interpolation. Image interpolation can be used in changing the size of a digital image to improve its appearance when viewed on a display device A sequence of image frames can also be interpolated along the temporal dimension.A 24 frames/sec motion picture can be converted to a

60 field/sec NTSC signal for TV through interpolation Temporal interpolation can also be used to improve the appearance of slow motion video.

Interpolation can also be used in other applications such as image coding.For example, a simple approach to bit rate reduction would be discard some pixels or some frames and recreate them from the coded pixels or frames.

2.6.1 Spatial Interpolation

Consider a 2-D sequence $f(n_1, n_2)$ obtained by sampling analog signal $f_c(x, y)$.

$$f(n_1, n_2) = f_c(x, y) \mid_{x=n_1T_1, n_2T_2} \tag{2.13}$$

$$and \;\; f_c(x, y) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f(n_1, n_2) h(x - n_1T_1, y - n_2T_2) \tag{2.14}$$

where h(x,y) is the impulse response of an ideal separable analog low-pass filter.

There are several difficulties in using equation 2.14 for image interpolation. The analog image $f_c(x, y)$ even with an antialiasing filter, is not truely band-limited, so aliasing occurs when $f_c(x, y)$ is sampled. In addition h(x,y) is an infinite extent function, so evaluation of $f_c(x, y)$ using equation 2.14 cannot be carried out in practice. To approximate the interpolation, one approach is to use a low pass filter h(x,y) that is spatially limited. For a spatially limited h(x,y) the summation in equation 2.14 has a finite numbers of nonzero terms. If h(x,y)is a rectangular window function given by

$$h(x, y) = 1, \quad -(T_1/2) \leq x \leq (T_1/2) \quad -(T_2/2) \leq y \leq (T_2/2) \tag{2.15}$$

then it is called a zero-order interpolation.In zero-order interpolation, $\hat{f}_c(x, y)$ is chosen as $f(n_1, n_2)$ at the pixel closest to (x,y). Other examples of h(x,y) which are more commonly used are functions of smoother shape such as the spatially limited Gaussian function.

Another simple method widely used in practice is bilinear interpolation, in this method,$f_a(x,y)$ is evaluated by the linear combination of $f(n_1,n_2)$ at the four closest pixels. The interpolated $\hat{f}_a(x,y)$ in the bilinear interpolation method is:

$$\hat{f}_a(x,y) = (1-\Delta_x)(1-\Delta_y)f(n_1,n_2) \tag{2.16}$$

$$+(1-\Delta_x)\,\Delta_y\,f(n_1,n_2+1)$$

$$+\,\Delta_x\,(1-\Delta_y)f(n_1+1,n_2)$$

$$+\,\Delta_x\,\Delta_y f(n_1+1,n_2+1)$$

$$where \quad \Delta_x = (x - n_1T_1)/T_1 \tag{2.17}$$

$$and \quad \Delta_y = (y - n_2T_2)/T_2 \tag{2.18}$$

spatial interpolation scheme can also be developed using motion estimation algorithms.One example,where an image frame that consists of two image fields is constructed from from a single image field is discussed in chapter(4).

Chapter 3

WAVELET TRANSFORM

3.1 Introduction

An important problem in signal processing is to define a representation that is well adapted for extracting the information content of signals. The sharp variations of signal amplitude are generally among the most meaningful features.For example the discontinuities of the image intensity provide the contours of the different objects. When the signal includes important structures that belong to different scales it is often helpful to reorganise the signal information into a set of details components of varying size. The wavelet transform is the linear operation that decomposes a signal into components that appear at the different scales. This transform is based on the convolution of the signal with a dilated filter.

3.2 Wavelets: A Brief Review

The wavelet transform of a signal $f(t)$ is by definition its convolution with a wavelet w(t) dilated by a factor (a);

$$W_f(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} f(t)\omega\left(\frac{t-b}{a}\right)dt \tag{3.1}$$

$$W_f(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} F(\omega)W(a\omega)e^{j\omega b}d\omega \quad (3.2)$$

which is equivalent to filtering the signal $f(t)$ with the bandpass filter $W(a\omega)$, whose bandwidth changes according to the scale parameter a. Clearly , large scales correspond to narrow smoothing filters that represent a global view of the signal $f(t)$ and small scales correspond to wide filters that look into the details of $f(t)$ (i.e. high frequency components).

Wavelet expansion of the signal $f(t)$ is essentially a decomposition of its frequency content using filters of constant relative bandwidth. The signal $f(t)$ can be recovered from its redundant wavelet transform coefficients using:

$$f(t) = \frac{1}{a^{5/2}} \int_a \int_b W_f(a,b)\omega\left(\frac{t-b}{a}\right) dadb \quad (3.3)$$

assuming that

$$\int_t W(t)dt = 0 \quad (3.4)$$

$$\int_\omega \left(\frac{W(w)^2}{\omega}\right) d\omega < \infty \quad (3.5)$$

As it is the case with the Fourier transform, where a signal is expanded in terms of of complex exponentials of different frequencies, a wavelet expansion involves dilations of a single wavelet ("mother wavelet"). The choice of a mother wavelet depends on the application, where a particular wavelet is chosen based on its time and frequency localisations.

Orthogonality is an important element of wavelet analysis, and a mother wavelet is orthogonal to its own dilations and translations. Wavelets provide orthonormal basis for expansions of functions that are not of single frequency and are therefore ideal for characterising signals with discontinuities.

The wavelet transform parameters can be discretised so that

$$C_{m,n} = a_0^{-m/2} \int_t f(t) w \left(\frac{t - n a_0^m T}{a_0^m} \right) dt \tag{3.6}$$

$$a = a_0^m \tag{3.7}$$

$$b = n a_0^m T \tag{3.8}$$

and T is the sampling period. The signal $f(t)$ can be recovered from its expansion coefficients using

$$f(t) = A \sum_m \sum_n C_{m,n} w \left(\frac{t - n a_0^m T}{a_0^m} \right) \tag{3.9}$$

where A is constant.

The case where $a_0 = 2$ is known as the dyadic wavelet transform, where the signal $f(t)$ is band-pass filtered using octave band filters This type of wavelet has the form

$$\psi_{m,n}(k) = 2^{-m/2} \psi (2^{-m} k - n) \tag{3.10}$$

$$m, n \in Z$$

The discrete wavelet transform (DWT) of discrete time sequence $f(k)$ is essentially a multiresolution characterisation of $f(k)$. Generally we take the DWT of a signal that is both time limited and resolution limited. A continuous time signal , uniformly sampled satisfies this criterion, [8]. A dyadic discrete wavelet transform is essentially a decomposition of the spectrum of $f(k)$; $F(w)$ into orthogonal subbands defined by

$$\frac{1}{2^j T} \leq w \leq \frac{1}{2^{(j+1)} T} \tag{3.11}$$

$$j = 1, 2, \ldots\ldots, J$$

where T is the sampling period associated with $f(k)$.

3.3 Zero Crossings of Wavelet Transform

If the wavelet function '$\psi_s(x)$' is a second derivative of a smoothing function '$\theta_s(x)$' , then the zero crossings of the wavelet transform $W_s(x)$ correspond to the local variation points of the signal f(x). In an image signal let I(x,y) be the representation of the image and '$\theta(x,y)$' be the smoothing function, which is used to smooth the image. In the case of separable smoothing function

$\theta(x,y) = \theta(x)\,\theta(y)$

Two wavelet functions are given by:

$$\psi_{s,H}(x,y) = s^2\frac{\partial^2\theta_s(x,y)}{\partial x^2} = \psi_s(x)\theta_s(y) \qquad (3.12)$$

$$\psi_{s,V}(x,y) = s^2\frac{\partial^2\theta_s(x,y)}{\partial y^2} = \psi_s(y)\theta_s(x) \qquad (3.13)$$

The two wavelet transform are given for horizontal and vertical direction as

$$W_{s,H}(x,y) = I * \psi_{s,H}(x,y) \qquad (3.14)$$

$$W_{s,V}(x,y) = I * \psi_{s,V}(x,y) \qquad (3.15)$$

If the smoothing function is Gaussian, than the detection of of zero crossing points of wavelet transform is equivalent to extraction of edges with Laplacian of Gaussian.

The DWT is implemented using a bank of bandpass and ~~low~~ high pass discrete time filters, h and g . As the input sequence $f(k)$ propagates through the filter bank tree of low pass and high pass filters , the output of the high pass filter g at stage m represents the detail signal which is a sampled version of the wavelet transform of $f(k)$ at the scale 2^m. At each stage the bandwidth of both filters is halved with the upper half band associated with the high pass filter g and lower half band associated

with the low pass filter h. The resulting signal decomposition $\{d^1, d^2, ..., d^m, c^m\}$ is the DWT of $f(k)$ i.e. c_0. Resolution level m is included for the sake of completeness. The filters thus each perform a convolution and decimation on input The input signal $c_0 = f(k)$ is called resolution level zero. Resolution levels $m-1$ and m are c^{m-1} and c^m respectively, and d^m is the difference level m It is called difference level because it represents the difference in the signal between c^{m-1} and c^m. Since DWT so implemented is a linear system, a normal process at input results in a normal process at output.

Finally due to discretisation of the wavelet transform parameter b, the wavelet expansion coefficients are no longer shift invariant. The expansion coefficients of a signal $f(t)$ may differ from those of $f(t - t0)$. Indeed the wavelet coefficients of a particular pattern are modified when the position of this pattern is changed.On contrary,it is clear that the position of the zero-crossings of a dyadic wavelet transform are translated when the signal $f(x)$ is translated.This property of the zero crossings, if the motion has taken place,is used for detecting motion of the objects

Chapter 4

MOTION ESTIMATION WITH ZERO CROSSINGS OF WAVELET TRANSFORM

4.1 Introduction

In order to accurately estimate motion many methods have been proposed and tested.These methods can be classified as Matching and Spatio-temporal gradient based methods. Although conventional block matching methods can estimate motion using relatively simple calculations, they are inadequate in some cases. For example, if block boundaries and the object boundaries do not coincide. In such cases the block size of the two cosecutive frame should be so chosen such that it contains the object. The Block matching method considers a small region in a frame and searches for the displacement which produces the best match among the possible regions in an adjacent frame. In this method the displacement vector $[dx, dy]$ is estimated by using the equations (2.9) and (2.10) given in the chapter (2).

The gradient based method works quite well if the magnitude of the motion is not large. To estimate motion parameters, which takes care of this problem in gradient based method, a new hierarchial motion estimation approach is proposed. It is

shown that when motion parameters are estimated hierarchially using gradient based method at each resolution level, it gives a better estimation of motion parameters than the Laplacian of Gaussian Method which is suitable only for small displacement per frame. A similar hierarchial method for motion estimation was used by (Cheong and Aizawa) [7]. They used the block matching technique for motion estimation. However interdependency of layers is a serious drawback for estimating motion. Here interdependency of layer means that the motion parameters calculated at the coarser layer is used to minimize the error expression at the next detail layer for estimating the motion parameters. In this way the motion vectors are hierarchially estimated. However if the motion estimation at the coarser layer is incorrect it is almost impossible to overcome the misleading estimation in the next detail layers.

If the wavelet function '$\psi_s(x)$' is chosen as the second derivative of a smoothing function '$\theta_s(x)$' , then the zero crossings of the wavelet transform $W_s(x)$ correspond to the local variation points of the signal f(x). In an image signal let I(x,y) be the representation of the image and '$\theta(x,y)$' be the smoothing function, which is used to smooth the image. For separable smoothing function

$\theta(x,y) = \theta(x)\ \theta(y)$

Two wavelet functions are obtained as:

$$\psi_{s,H}(x,y) = s^2\frac{\partial^2\theta_s(x,y)}{\partial x^2} = \psi_s(x)\theta_s(y) \tag{4.1}$$

$$\psi_{s,V}(x,y) = s^2\frac{\partial^2\theta_s(x,y)}{\partial y^2} = \psi_s(y)\theta_s(x) \tag{4.2}$$

The wavelet transforms for horizontal and vertical direction are given by:

$$W_{s,H}(x,y) = I * \psi_{s,H}(x,y) \tag{4.3}$$

$$W_{s,V}(x,y) - I * \psi_{s,V}(x,y) \tag{4.4}$$

The first step of motion estimation is, segmentation of the image into stationary and moving areas Once the stationary pixels are known they can be ignored in the following stage of motion estimation. Segmentation is obtained by applying the simple thresholding procedure

Moving If

$$| I_1(x,y) - I_2(x,y) | \quad > T_1 \tag{4.5}$$

Stationary If

$$| I_1(x,y) - I_2(x,y) | \quad \leq T_1 \tag{4.6}$$

Where 'T' is the threshold chosen by the user, depending upon noise, image intensities etc.

On evaluating the zero crossings of wavelet transform ; the difference of zero crossing will be non zero if and only if the image has moved. Although the wavelet coefficients of a particular pattern are modified when the position of this pattern are changed, Yet the position of the zero crossing of a wavelet transform are translated when the signal f(x) is translated. The extracted zero-crossing points represents the features of the image, but these points seldom characterise the objects boundaries completely. But here we are only interested in the zero crossings.

4.2 Hierarchial Motion Estimation

For a given image sequence, multiresolution image decomposition is obtained by using a biorthogonal wavelet transform. The dydic wavelet '$\psi(x)$' used for the wavelet transform and the corresponding impulse response coefficients of filters 'H' and 'G' corresponding to the wavelet is shown in fig (4 1) and Table (4 1).

The magnitude of the motion vector computed at coarser layer in the multiresolution representation are used to initialise the estimation of magnitude at the finer layer. Let 'D_j' be the magnitude of the average value of the motion vector computed at resolution 2^{-j}.In the next resolution $2^{-(j+1)}$, the D_j is taken to initialise the calculation of motion vector magnitude 'D_{j+1}'. At the next resolution that is $2^{-(j+2)}$, D_j and D_{j+1} both the values are used and the value of magnitude of motion vector calculated at each pixel are compared with both D_j and D_{j+1}. Finally at the finest layer the magnitude vector is calculated by comparing with the motion vector magnitudes of the two previous layers. The angle 'θ' is calculated at the finest layer.

The steps involved in the motion estimation can be summerised as under:

1. First the wavelet transform of the image sequence frame is taken. For each image frame, the intensity 'I' is convolved with second derivative of smoothing function or the wavelet function.

2. The wavelet transform decomposes the image into different resolution levels containing coarser to detail informations as the resolution level is increased. The image of the same resolution level is separated out in each frames for evaluation of motion parameters hierarchially.

3. The estimation of motion vector D_j at coarser layer 2^{-j}, obtained as follows:

    (a) The image sequences at the coarser layer 2^{-j} in two successive frames is taken.

    (b) The temporal difference of Image sequence at this resolution step is calculated.

(c) Gradient of image dI/dx and dI/dy is calculated using a'3*3' Sobel operator.

(d) Calculation of magnitude of motion vector and direction of motion at zero crossing is calculated by:

$$U_n = \frac{\frac{\partial I}{\partial t}}{\sqrt{(\frac{\partial I}{\partial x})^2 + (\frac{\partial I}{\partial y})^2}} \tag{4.7}$$

and the angle 'θ' between the unit normal to the contour and the positive x-axis is given by

$$\theta = tan^{-1}\frac{\frac{\partial I}{\partial y}}{\frac{\partial I}{\partial x}} \tag{4.8}$$

(e) Finally the values of U_n and angle 'θ' is averaged out for which are not zero.

4. At resolution $2^{-(j+1)}$,that is the next higher resolution; the above procedure is repeated. But for magnitude calculation in (d) above, each value of magnitude is compared with D_j and all values less than D_j are discarded. Then average of all these values thus calculated is taken. Let this be D_{j+1}.

5. At the next layer $2^{-(j+2)}$, the magnitude of the motion is compared with D_j and D_{j+1} simultaneously.That is,all magnitude values which are greater than D_{j+1} are kept,and if no values are greater than D_{j+1} than values greater than D_j are kept and finally average of all these values thus calculated is taken.

6. In a similar way the magnitude of motion vector at the finest layer is calculated.

7. Similarly, Direction vector is calculated at the finest layer as calculated in 3(d) and average is taken of all the values so calculated which are not zero.

The motion vectors were also calculated using Laplacian of Gaussian smoothing function.The steps involved are as follows:

1. The image intensity 'I' is first convolved with Laplacian of Gaussian smoothing function.
2. The temporal difference of Image sequence is calculated.
3. Gradient of image dI/dx and dI/dy is calculated using a '3*3' Sobel operator.
4. Calculation of magnitude and direction of motion at zero crossing is carried out by

$$U_n = -\frac{\frac{\partial I}{\partial t}}{\sqrt{(\frac{\partial I}{\partial x})^2 + (\frac{\partial I}{\partial y})^2}} \tag{4.9}$$

   and the angle 'θ' between the unit normal to the contour and the positive x-axis is given by

$$\theta = tan^{-1}\frac{\frac{\partial I}{\partial y}}{\frac{\partial I}{\partial x}} \tag{4.10}$$

5. Finally the values of U_n and angle 'θ' is averaged out for the values which are not zero.

4.3 Experiment Results and discussion

4.3.1 Synthetic Image Sequence

A Moving Square.A synthetic image sequence used in this experiment consisted of a bright square on a dark background.The total image size as 256*256 pixels. The square started in the upper left corner and moved diagonally towards the lower right corner by 1.414 pixel/frame fig (2). Magnitude and direction are calculated using the

Laplacian of Gaussian and also by multi-resolution decomposition of image sequence frames using wavelet transform of sequence frames. The observations are:

1. Laplacian of Gaussian.

Magnitude			Direction in Degrees		
Actual	Calculated	Error	Actual	Calculated	Error
(a) First and Second Frame.					
1.414	0.7	0.7	135	159	24
(b) First and Third Frame.					
2.828	1.4	1.42	135	159	24
(c) Second and Third Frame					
1.414	0.7	0.7	135	159.16	24.16

2. Wavelet Transform of Image Sequence.

Magnitude			Direction in Degrees		
Actual	Calculated	Error	Actual	Calculated	Error
(a) First and Second Frame.					
1.414	1.422	0.01	135	134	01
(b) First and Third Frame.					
2.828	2.862	0.04	135	134	01
(c) Second and Third Frame.					
1.414	1.457	0 04	135	138	03

4.3.2 Laboratory Image Sequence

Because of the limitation in equipment it was not possible to obtain real time image sequences of natural scene. Instead a stationary camera was used to view a model of moving circle. This circle even contains fine scale structures fig (4.2). Two such frames were taken for estimation of motion parameters. The followings are the observations:

1. Laplacian of Gaussian.

Magnitude			Direction in Degrees		
Actual	Calculated	Error	Actual	Calculated	Error
(a) First and Second Frame.					
12.69	3.3	9.39	18.43	23.95	05.52

2. Wavelet Transform Image Sequence.

Magnitude			Direction in Degrees		
Actual	Calculated	Error	Actual	Calculated	Error
(a) First and Second Frame.					
12.69	12.622	0.07	18.43	17.80	00.63

The followings are the observations:

1. Laplacian of Gaussian can estimate only small values of motion vector, and failed to estimate larger values of motion vectors.

2. Error in estimated values of motion parameters ; when Laplacian of Gaussian method was used were quite large, both for magnitude and direction vectors. Whereas the hierarchial wavelet transform method worked quite well in estimating small as well as larger values of motion vectors.

3. In normal gradient based method the values of dI/dx and dI/dy is zero or very small in homogeneous region, it is difficult to estimate the velocity there. Therefor the pixels whose gradient value $| dI/dx + dI/dy |$ is greater than a particular threshold is taken into consideration for calculations. That is manual thresholding is required. But in hierarchial method of wavelet transform this is taken care of automatically through initialisation as we move from coarser layer to the next layer.

4. The number of frames required for reasonable correct estimation of motion vectors in the case of Laplacian of Gaussian are quite large in numbers. Whereas in multi- resolution wavelet transform pseudo frames are created as one image frame is decomposed into various layers of image and reasonably accurate estimate can be carried out using just two frames. This is shown experimentally.

Here we see that since the pixel displacement per frame was large the conventional gradient based method using Laplacian of Gaussian as a smoothing function failed to estimate the motion vector parameters. Whereas multi resolution of zero crossings of wavelet transform could estimate the motion parameters quite accurately.

4.3.3 Application of Motion Estimation Method to Spatial Interpolation

The spatial interpolation technique may be used in creating a frame from a field,but incorporating some additional knowledge about image may improve the performance of a spatial interpolation algorithms.Let $f(x, y_{1})$ and $f(x, y_0)$ denote image intensities of two adjacent horizontal scan lines of a field.We wish to create a new horizontal scan line between $f(x, y_{1})$ and $f(x, y_0)$. One model that takes into account the

spatial continuity of such elements as contours and scratches is.

$$f(x, y_0) = f(x - d_x, y_{-1}) \tag{4.11}$$

Where d_x is the horizontal displacement between y_{-1} and y_0. Equation 4.11 can be viewed as a spacial case of uniform translational velocity model equation 2.1 The spatial variable 'y' in equation 4.11 has a function very similar to the time variable 't' as in equation 2.1 and there is only one spatial variable 'x' in 4.11,while there are two spatial variable 'x' and 'y' in equation 2.1. As a result the problem is now of estimating d_x in equation 4.11. For example under the assumption of uniform velocity, equation 4.11 can be expressed as:

$$f(x, y) = f(x - V_x(y - y_{-1}), y_{-1}) \quad y_{-1} < y \leq y_0 \tag{4.12}$$

$$\textit{which leads to} \quad V_x \frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y)} = 0 \tag{4.13}$$

Thus once the motion parameters are estimated these can be used for spatial interpolation. This is shown in fig (4.10) and fig (4.11). The new frame is created with the knowledge of motion parameters

Table 4.1. Wavelet Filter Coefficients Used for Calculations

	H	G
0	0.4347	0.7118
1	0.2864	-0.2309
2	0.0450	-0.1120
3	-0.0393	-0.0226
4	-0.0132	0.0062
5	0.0032	0.0039

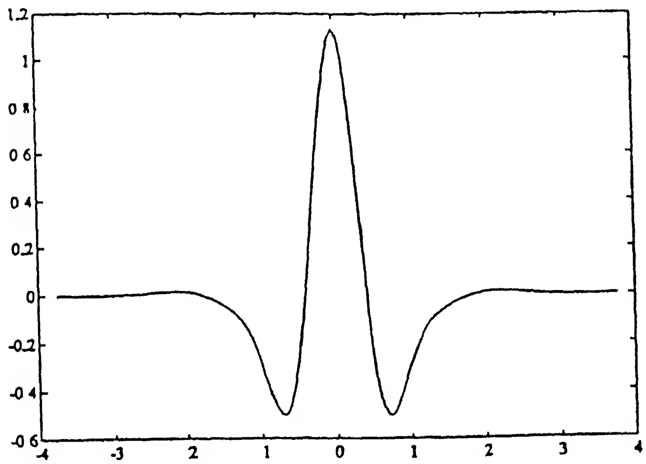

Figure 4.1: Graph of the Dyadic Wavelet $\psi(x)$ used in the Numerical Experiment

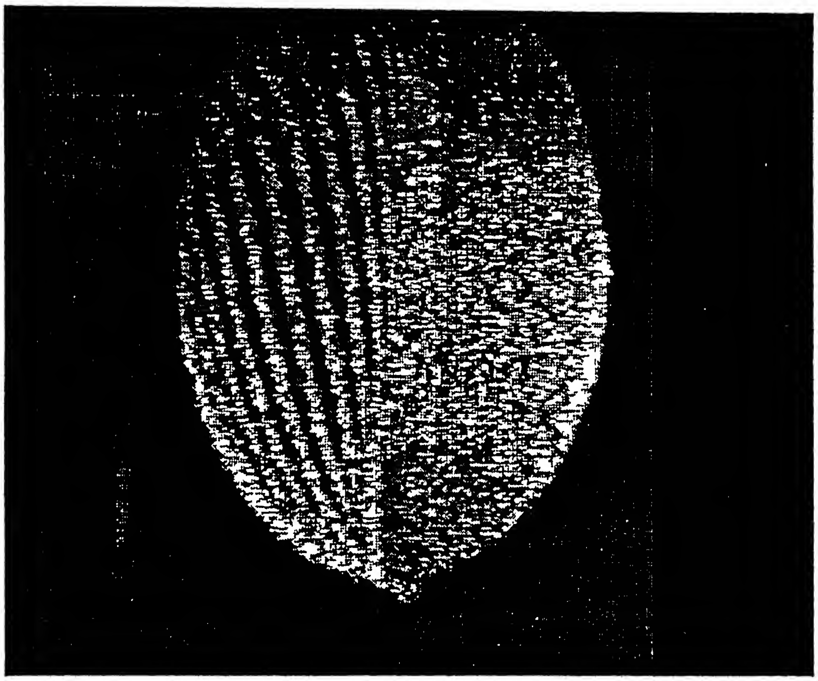

Figure 4.2: Test Image of Circle

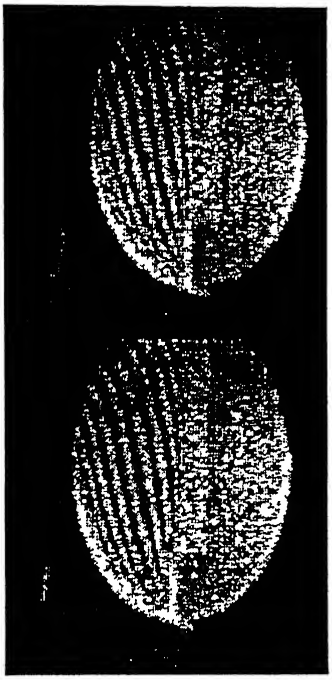

Figure 4.3: Images of Circle Two Frames Side by Side

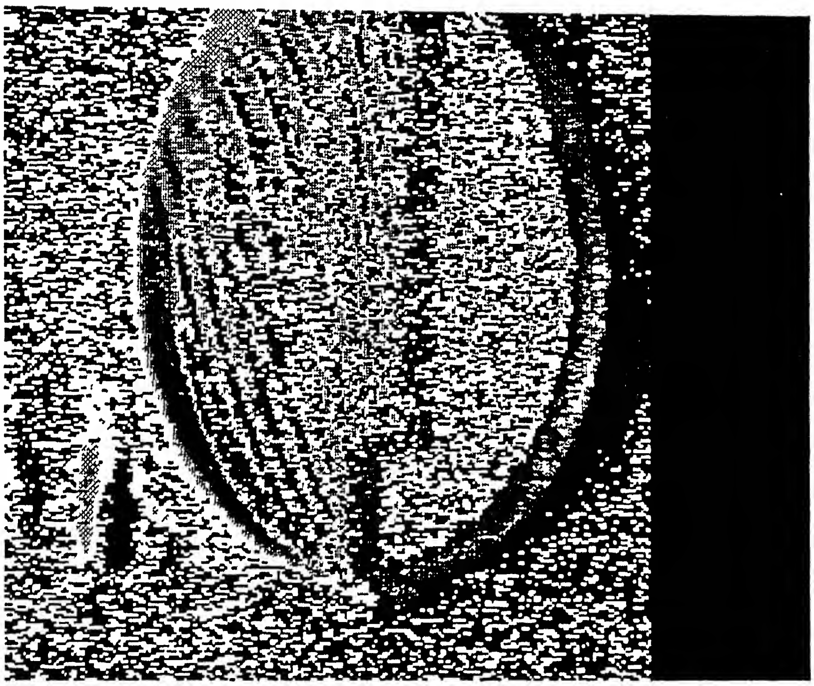

Figure 4.4. Superimposed Frames showing Dislacement of Circle

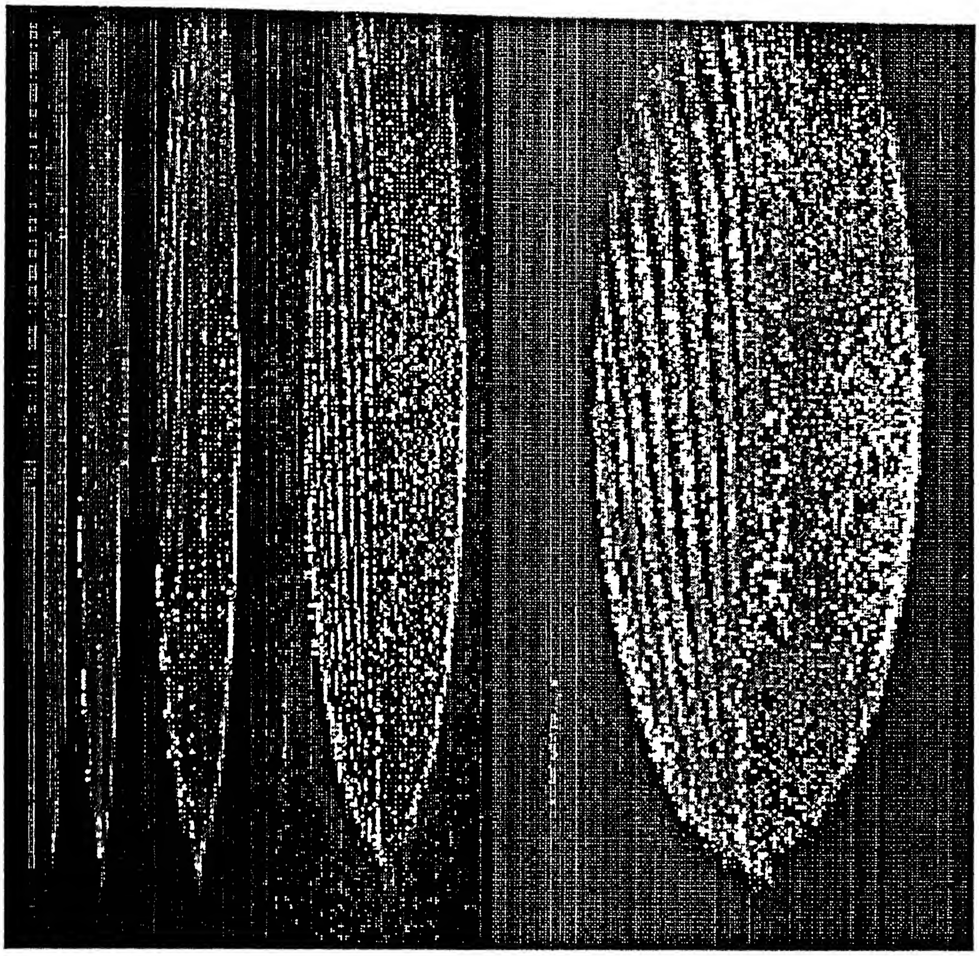

Figure 4.5: Wavelet Transform of Circle Image at Different Resolution

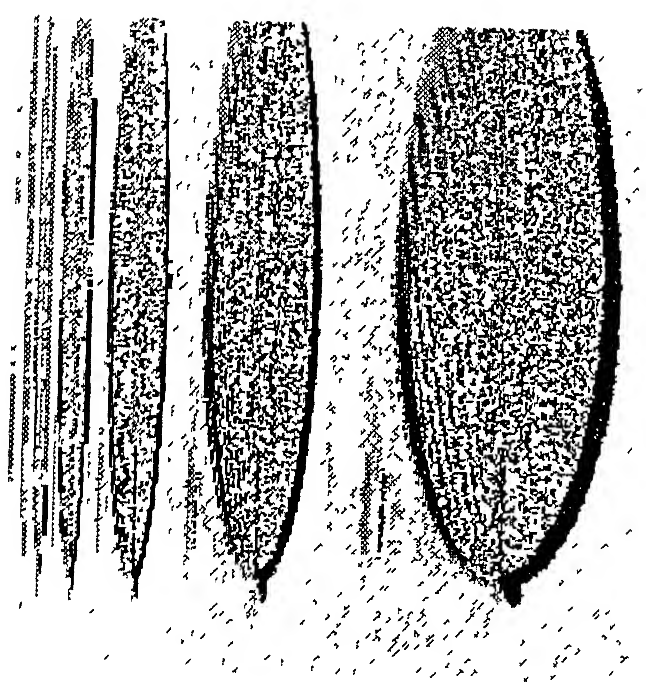

Figure 4.6: Wavelet Transform Zero Crossings of Circle showing Displacement at Different Resolution

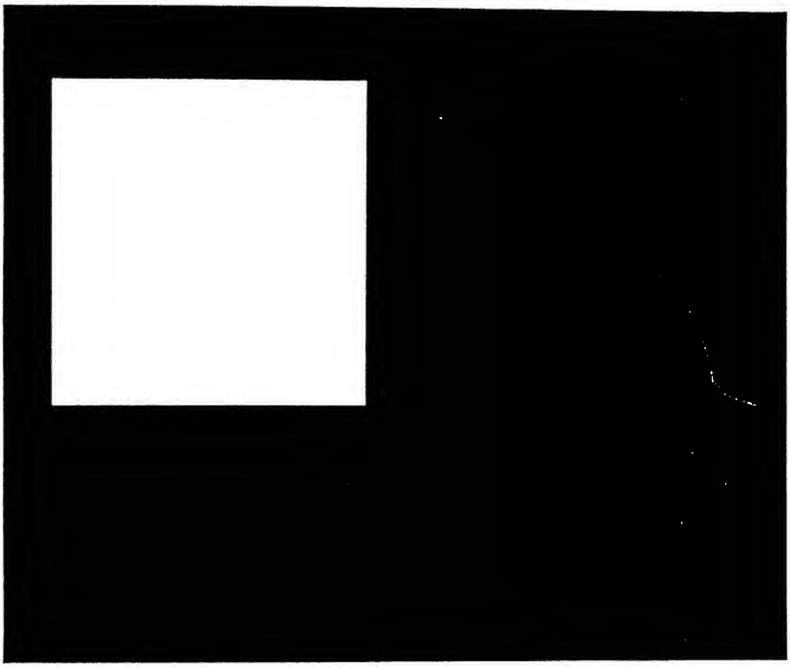

Figure 4.7: Image of Test Figure Square

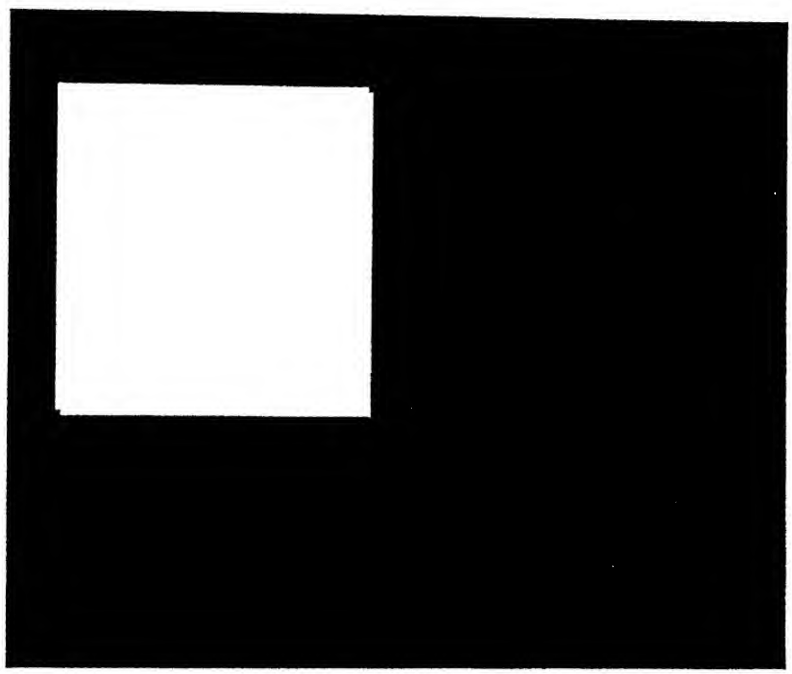

Figure 4.8. Image of the Superimposed Frames of Square showing Displacement

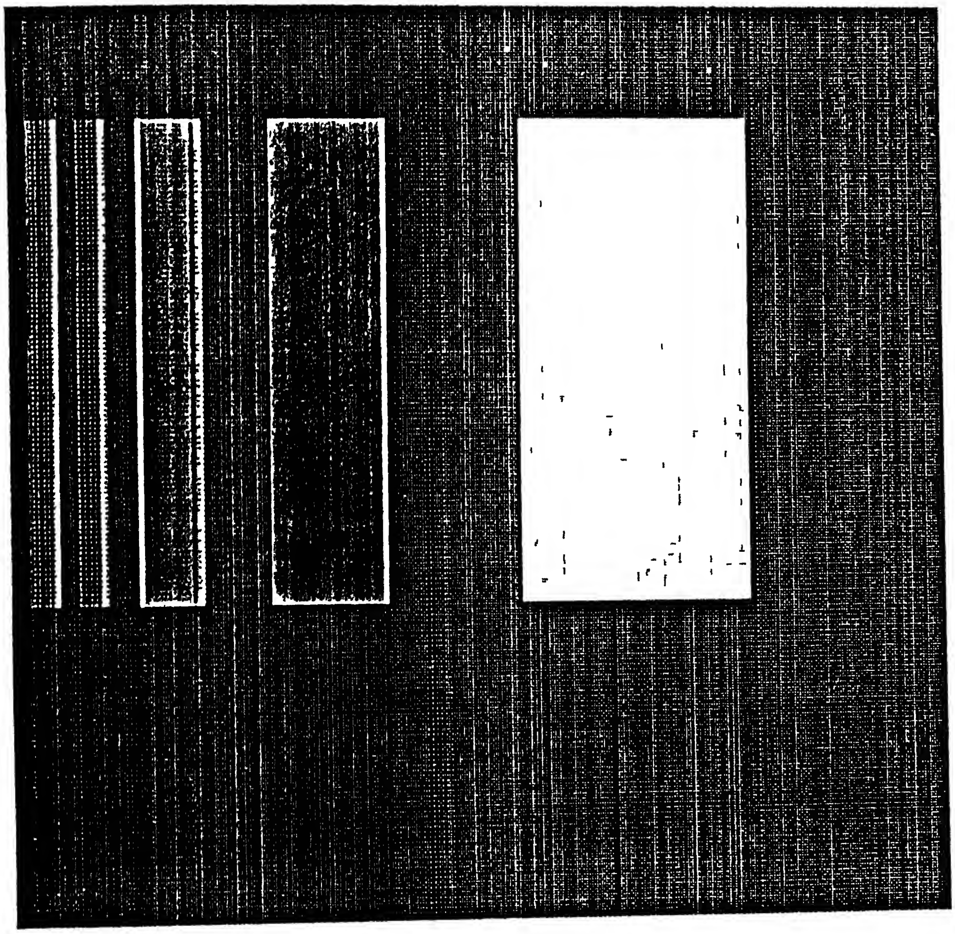

Figure 4.9: Wavelet Transform Image of Square at Different Resolution

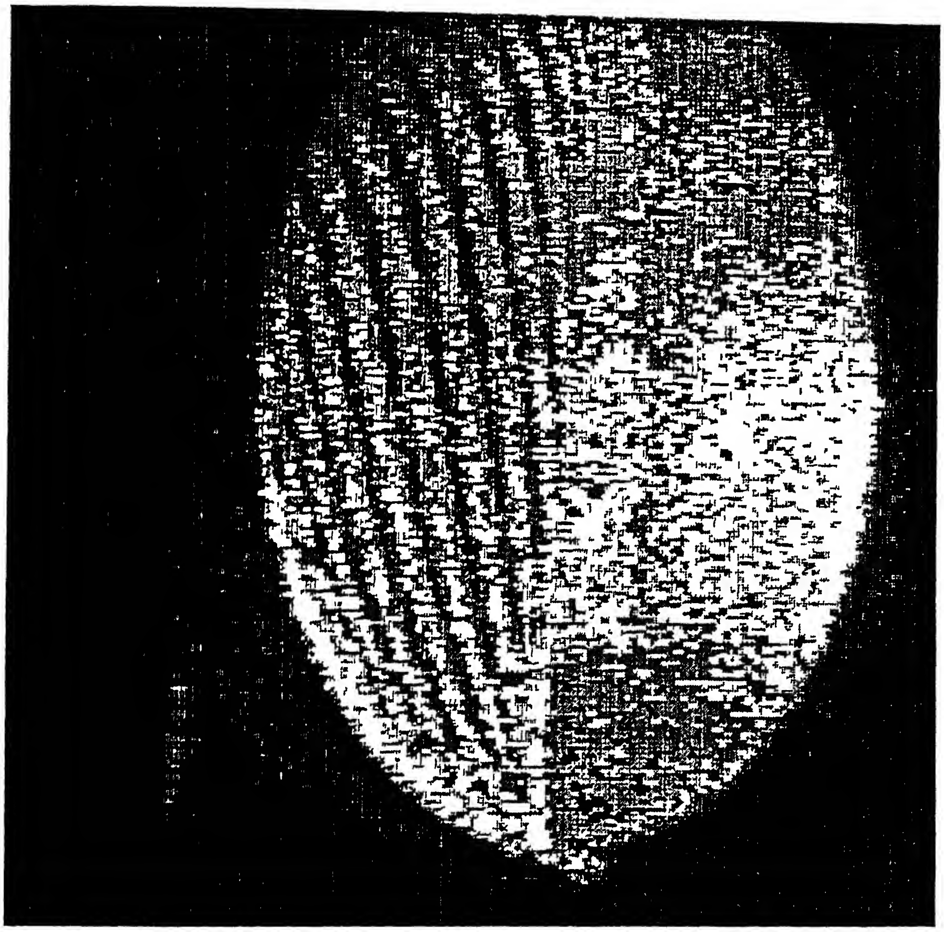

Figure 4.10: Reconstructed Frame of Circle

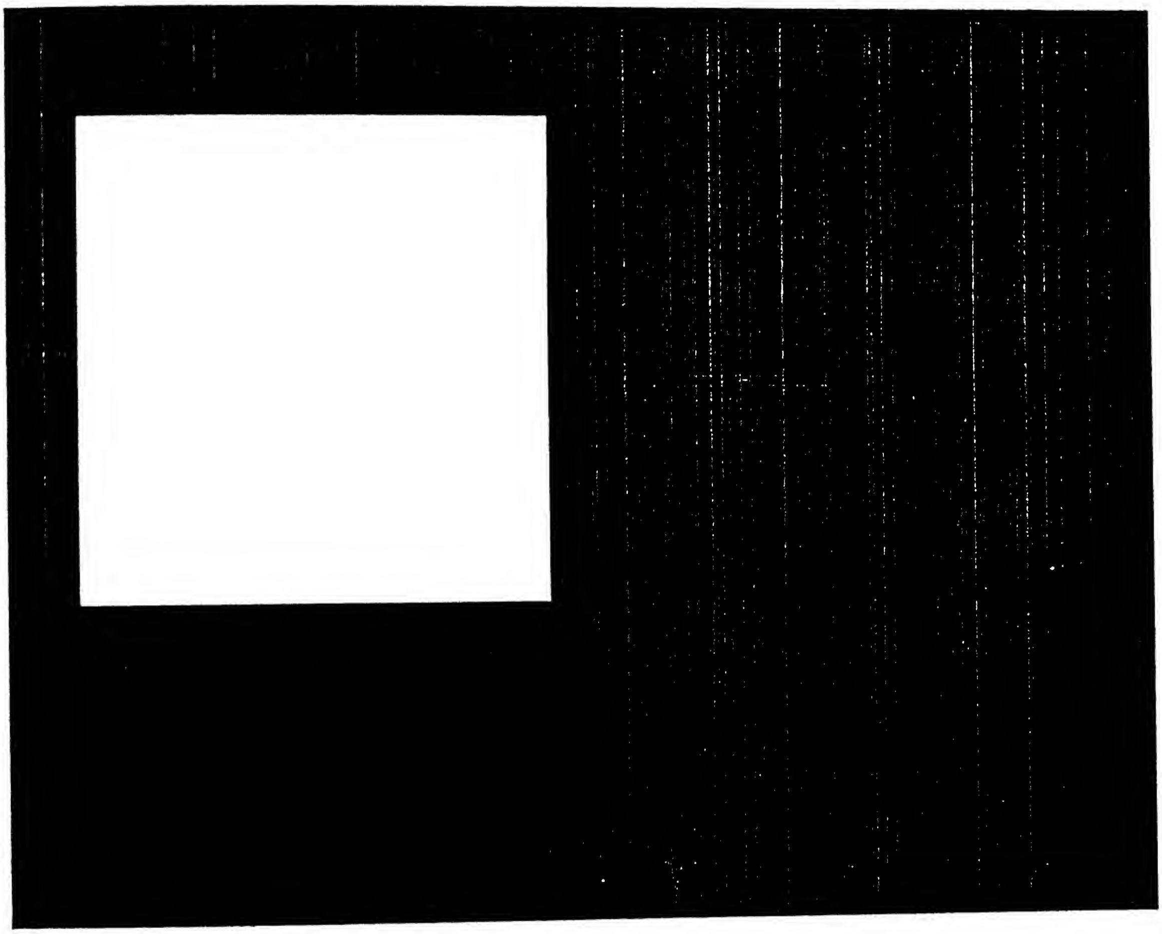

Figure 4.11: Reconstructed Frame of Square

Chapter 5

CONCLUSION

In this work the hierarchial method of motion estimation using zero crossings of wavelet transform is proposed. The gradient based constrain equation is used for motion estimation.It is shown that this method eliminates the drawbacks of block matching method as well as that of the conventional gradient method.

Block matching method fails if the initial estimation is wrong and is computationally expensive. The conventional gradient based method using Laplacian of Gaussian smoothing function works well for small values of pixel displacement per frame. Whereas the hierarchial method presented can, not only estimate motion vectors accurately but is suitable for large values of pixel displacement per frame.

5.1 Scope of Future Work

The scope of future work in this area is plenty First of all the method can be extended to cater for zooming and rotationary movements of objects.It can also be extended to global and local detection of motion with various objects moving with different velocities.

References

[1] J.K.Aggarwal and N.Nandhakumar. On the computation of motion from sequences of images. *Proc.IEEE.*, vol-76: pp.917 935, Aug. 1988.

[2] B.F.Buxton and H.Buxton . Computation of optical flow from the motion of edge features in image sequences. *Image Vision Comput.*, vol.2 no. 11: pp.59–75, - 1984.

[3] James H. Duncan and Tsai Chia Chou. On detection of motion and the computation of optical flow. *IEEE Trans. on PAMI*, vol. 14 no. 3: pp.346 352, Mar. 1992.

[4] J.K.Aggarwal R.Jain and W.N.Martin. Segmentation through detection of changes due to motion. *Comput. Vision Graphics Image Processing*, vol. 11: pp.13 34, -. 1979.

[5] Stephane Mallet. Zero crossings of a wavelet transform. *IEEE Trans. on IT*, vol-17 No. 4: pp.1019 1033, Jul. 1991.

[6] P.Lilov and L.Tsekov. Motion estimation using image histogram sequence analysis. *Signal Processing .*, vol-30 No.3: pp.373 384, Feb. 1993.

[7] Cha Keon Cheong and K.Aizawa. Motion Estimation with Wavelet Transform and the Application to Motion Compensated Interpolation. *ICASSP* ., pp.V 217-V220, 1993.

[8] Olivier Rioul and Martin Vetterli. Wavelets and signal processing. *IEEE SP Magazine*, pages pp.14-38, Oct. 1991.